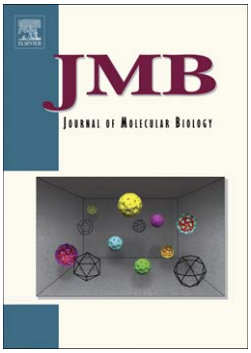
A combined NMR and computational approach to investigate peptide binding to a designed armadillo repeat protein

Ewald, Christina ; Christen, Martin T ; Watson, Randall P ; Mihajlovic, Maja ; Zhou, Ting ; Honegger, Annemarie ; Plückthun, Andreas ; Caflisch, Amedeo ; Zerbe, Oliver

Abstract: The specific recognition of peptide sequences by proteins plays an important role both in biology and in diagnostic applications. Here we characterize the relatively weak binding of the peptide neurotensin (NT) to the previously developed Armadillo repeat protein $VG_3$28*byamultidisciplinaryapproachbasedonsolutionNMRsp*
sizebinderwithverysimilaraffinityforNT, *forwhichnear−completebackboneassignmentswereachieved.Abindingmode*
labeledNT.FavorableintermolecularinteractionsareobservedintheMDsimulationsfortheresiduesthatwerepreviouslysi
capforproteinstabilityandpeptidebinding.Ourmultidisciplinaryapproachdemonstratesthataninitiallow−
resolutionpictureforalow−micromolar−peptidebindercanberefinedthroughthecombinationofNMR, *proteindesign*, *do*

DOI: https://doi.org/10.1016/j.jmb.2015.02.022



A Combined NMR and Computational Approach to Investigate Peptide Binding to a Designed Armadillo Repeat Protein

Christina Ewald, Martin T. Christen, Randall P. Watson, Maja Mihajlovic, Ting Zhou, Annemarie Honegger, Andreas Plückthun, Amedeo Caflisch, Oliver Zerbe

PII: S0022-2836(15)00197-7
DOI: doi: 10.1016/j.jmb.2015.02.022
Reference: YJMBI 64720

To appear in: *Journal of Molecular Biology*

Received date: 3 December 2014
Revised date: 18 February 2015
Accepted date: 23 February 2015

Please cite this article as: Ewald, C., Christen, M.T., Watson, R.P., Mihajlovic, M., Zhou, T., Honegger, A., Plückthun, A., Caflisch, A. & Zerbe, O., A Combined NMR and Computational Approach to Investigate Peptide Binding to a Designed Armadillo Repeat Protein, *Journal of Molecular Biology* (2015), doi: 10.1016/j.jmb.2015.02.022

A Combined NMR and Computational Approach to Investigate Peptide Binding to a Designed Armadillo Repeat Protein

Christina Ewald[1], Martin T. Christen[1†], Randall P. Watson[1†], Maja Mihajlovic[2], Ting Zhou[2], Annemarie Honegger[2], Andreas Plückthun[2]*, Amedeo Caflisch[2]*, Oliver Zerbe[1]*

[1] Department of Chemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

[2] Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

*corresponding authors:

Email address of corresponding authors: oliver.zerbe@chem.uzh.ch; caflisch@bioc.uzh.ch; plueckthun@bioc.uzh.ch

[†] authors contributed equally

Abstract

The specific recognition of peptide sequences by proteins plays an important role both in biology and in diagnostic applications. Here we characterize the relatively weak binding of the peptide neurotensin (NT) to the previously developed Armadillo repeat protein (ArmRP) VG_328 by a multidisciplinary approach based on solution NMR spectroscopy, mutational studies, and molecular dynamics (MD) simulations, totaling 20 μs for all MD runs. We describe assignment challenges arising from the repetitive nature of the protein sequence, and present novel approaches to address them. Partial assignments obtained for VG_328 in combination with chemical shift perturbations (CSPs) allowed us to identify the repeats not involved in binding. Their subsequent elimination resulted in a reduced-size binder with very similar affinity for NT, for which near-complete backbone assignments were achieved. A binding mode suggested by automatic docking and further validated by explicit solvent MD simulations is consistent with paramagnetic relaxation enhancement (PRE) data collected using spin-labeled NT. Favorable intermolecular interactions are observed in the MD simulations for the residues that were previously shown to contribute to binding in an Ala scan of NT. We further characterized the role of residues within the N-cap for protein stability and peptide binding. Our multidisciplinary approach demonstrates that even in the absence of crystallographic data an initial low-resolution picture for a low μM peptide binder can be refined through the combination of NMR, protein design, docking and MD simulations to establish its binding mode, thereby providing valuable information for further design.

Abbreviations

ArmRP, Armadillo Repeat Protein; M^x, R^x, A_x and Y_x: consensus and randomized internal, as well as C- and N-capping repeats (numbered sequentially in Arabic superscript, version (*i.e*. generation) of the design in Roman subscript); CSP, chemical shift perturbations (Δδ in ppm); ΔΔδ, differential CSP; ELISA, enzyme-linked immunosorbent assay; HSQC, heteronuclear single-quantum coherence spectroscopy; IMAC, immobilized metal ion affinity chromatography; K_d, equilibrium dissociation constant; MD, molecular dynamics; MTSL, *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methane-sulfonothioate; NOE, nuclear Overhauser effect; NT, neurotensin; PRE, paramagnetic relaxation enhancement; RMSF, root mean square fluctuations; SEC, size-exclusion chromatography; SPR, surface plasmon resonance; VG_328, selected member of a randomized ArmRP library.

Introduction

Easy access to oligonucleotides of any desired sequence via DNA synthesis has revolutionized molecular biology, allowing manipulation of genetic material to become a routine task. On-demand availability of binding proteins that bind peptides or extended parts of target proteins in a sequence-specific manner could have a transforming effect on various fields such as proteomics, structural biology, medical diagnostics and even therapy. Many proteins of interest have disordered termini or loosely packed loops, however, no binding proteins have yet been developed to allow target binding in a rational way based on a target sequence.

Currently, monoclonal or recombinant antibodies [1] and a range of other scaffolds [2-6] are available as protein- or peptide-binding reagents. The most prominent drawback of all of these systems is that for each new target a completely new binder must be established and that previously established binders for similar targets do not provide sufficient design information for future projects. Furthermore, many of these scaffolds preferentially bind to the surface of *folded* proteins, but unfolded protein or peptides also play a vital role in cellular signaling and protein trafficking.

Antibodies bind unstructured peptides with high affinities [7], but their mode of binding is not conserved. Moreover, antibodies and their derivatives contain disulfide bonds, which do not form when expressed in the cytoplasm, rendering them

unsuitable for intracellular applications. In contrast, small adaptor domains like SH2, SH3 and PDZ domains [8] usually show specific binding in a conserved fashion. However, their binding affinity is weak, only short sequences are recognized, and specificity is limited to the recognition of a few sequence motifs. Repeat proteins, in particular Armadillo repeat proteins (ArmRPs) [9], tetratricopeptide repeats [10], WD40 proteins [11], HEAT repeats [12], and Ankyrin repeats [13], possess an intrinsic ability to bind peptides due to their repetitive structure, resulting in well-defined surfaces that can be used for binding. ArmRPs, which are abundant in eukaryotes [14], often mediate protein-protein interactions and participate in a broad range of biological processes [15]. Well-known examples are β-catenin, which is involved in cell adhesion and signaling [16], and importin-α, which is vital for the nucleocytoplasmic transport of proteins [17].

Repeat modules of ArmRPs typically contain about 42 amino acids [18], which are arranged into a triangle of three α-helices (H1-3). In nature, 4 to 12 repeats are stacked beside each other forming a right-handed superhelix, the armadillo domain, which is responsible for peptide recognition [19]. Specialized capping modules at the N- and C-termini protect the elongated hydrophobic core. Peptides are bound in an extended conformation via interactions between highly conserved asparagine side chains, located in a groove formed by the third helices (H3) of each repeat, and the peptide backbone [20]. Specificity is conferred by other residues on the surface of H3 interacting with side chains of the target peptide. Each repeat of the armadillo domain specifically recognizes a dipeptide subunit of the bound peptide, providing the basis for a modular approach. Dissociation constants (K_d) as low as 10-20 nM have been reported [21,22]. Designed ArmRPs based on natural consensus sequences are available from previous studies [23]. These designed ArmRP scaffold proteins are soluble, highly expressed, stable in an intracellular environment, monomeric and display improved biophysical characteristics compared to natural ArmRPs. The original design by Parmeggiani *et al*.,[24] based on a sequence consensus of the importin-α and the β-catenin families, was further improved by Alfarano *et al*. [25] using a molecular dynamics (MD) based approach. The resulting scaffold is very stable and was employed in the creation of randomized libraries by Varadamsetty *et al*. [23].

A member of this library, VG_328, in which residues on the surface of H3 of the central three repeats have been randomized, was selected to bind the human

neurotensin (NT) peptide using ribosome display. NT (sequence: QLYENKPRRPYIL) was chosen as the target peptide for its lack of defined structure in solution [26]. The selected 32 kDa ArmRP VG_328 binds NT with a K_d of 7 μM at 4° C and contains five internal repeats flanked by N- and C-terminal capping repeats. Using ELISA assays with single-site alanine mutants of NT, VG_328 has been shown to specifically bind NT with four key NT side chains $Pro7^{NT}$, $Arg8^{NT}$, $Arg9^{NT}$, and $Tyr11^{NT}$ contributing to the moderate binding (for clarity, all peptide residues are listed with a "NT" superscript).

In order to mature designed ArmRPs towards higher affinity, it was pivotal to determine the binding mode of the first generation binder VG_328 to NT. The aim of the present study was to analyze this interaction in detail to guide future design efforts. None of the NT binders based on VG_328 yielded crystals of sufficient quality for structure determination by X-ray crystallography. Therefore, NMR studies, that also allow the characterization of interactions with K_d values in the mM to μM range [27], were performed to determine the binding location and orientation of NT on VG_328, and to establish if NT binds in the canonical orientation observed in natural ArmRPs.

Unfortunately, the repetitive nature of the armadillo sequence results in a number of technical challenges, some of which were already encountered in our previous studies of Ankyrin repeat proteins [28]. To facilitate this process we employed a wide range of techniques, from isotopic labeling and fragmentation of ArmRPs, to the selective deletion of repeats. This "reductive engineering" culminated in the design of a reduced-size binder that was much more amenable to NMR analysis and shows similar affinity for the NT peptide. The backbone of this minimal binder could be assigned nearly completely, and subsequent chemical shift perturbation experiments in combination with PREs from ligand-attached spin labels allowed the derivation of experimental constraints for peptide binding analysis. In combination with docking and MD simulations, a picture of the complex at atomic resolution emerged that will be very useful in future design rounds.

Results

All NMR methods that are suitable to establish the binding mode of a peptide require at least backbone and usually also side chain chemical shift assignments of the binding protein. The task of characterising peptide binding to repeat proteins, however, required the development of new tools for assigning the spectra of proteins with highly repetitive amino acid sequences. The combined use of biophysical, biochemical and computational techniques in a tour-de-force helped to improve the properties of the binder and allowed us to gain insight into its folding properties.

In what follows we first describe how we improved the properties of the original binder VG_328 ($Y_IMR^1R^2R^3MA_I$) to make it amenable to detailed NMR studies. (For nomenclature see Experimental Procedures; for sequence details see also Supplementary Figure S1). We then report on the attempts to solve the assignment problem inherent to repeat proteins using N-terminally truncated versions. Knowledge from these protein truncations and from chemical shift perturbation (CSP) data from VG_328 was then employed, in turn, to design smaller binders that could be assigned to a large extent. Finally, we present mutagenesis data used to deconvolute contributions from individual N-cap residues to NT binding and use MD calculations to probe NT binding and the behavior of the N-cap.

Chemical shift assignments are usually performed using $^{15}N,^{13}C$-labeled proteins (as well as perdeuteration, depending on size) and triple-resonance experiments [29,30]. Follow-up work by Alfarano *et al*. on the original scaffold indicated that consensus ArmRPs could be significantly stabilized by introducing two mutations (originally described as Q240L, F241Q [25]) in the C-cap to form the A_{II}-cap. The resulting binder $Y_IMR^1R^2R^3MA_{II}$ displayed essentially identical peptide binding and was sufficiently stable in perdeuterated form for NMR studies. While the C-cap mutant retained binding of NT, all N-cap mutations known to improve stability [25] also abolished NT binding (data not shown). We therefore continued our spectroscopic studies using the original less stable N-cap and the stabilized C-cap in $Y_IMR^1R^2R^3MA_{II.}$

Two sets of triple-resonance experiments, using samples with and without the peptide, were recorded for assignment purposes following a strategy described by Wetzel *et al*. [28]. For more details of spectroscopy and assignment, see the Supplementary

Materials and Methods. Although spectra were generally of very good quality (see Figure S2), complete backbone assignments proved impossible due to peak overlap and degeneracy of ^{13}C chemical shifts between residues at identical positions in the different repeats. Due to the repetitive nature of the internal repeats, few signals outside of the C-cap and H3 of the R^3 repeat could be unambiguously assigned. Nevertheless, randomized positions in H3 are unique and hence provided valuable assignment anchors and allowed unambiguous assignment of some protein segments of the putative peptide-binding surface. However, in the H1 and H2 helices (non-randomized) many assignment fragments could not be unambiguously mapped onto the sequence. Furthermore, in the absence of bound peptide, signals from residues of the N-cap were missing completely, most likely due to conformational exchange. Interestingly, some of these missing N-cap signals were observed in the spectra of the complex formed with NT. In the absence (and presence) of NT, 20.6% (25.4%) of backbone resonances of all non-Pro residues were assigned overall, including 92.7% (92.7%) of the C-cap and 94.7% (100%) of H3 of the R^3-repeat. [^{15}N, ^{1}H]-HSQC-based chemical shift mapping experiments of $Y_IMR^1R^2R^3MA_{II}$ with NT revealed that, in agreement with its moderate $K_d \approx 19 \pm 8$ μM, at 32 °C the system is in fast exchange.

Truncation of $Y_IMR^1R^2R^3MA_{II}$ Aids in Backbone Assignments

To reduce chemical shift degeneracy we investigated whether it was possible to truncate the protein by one repeat at a time, ultimately allowing deconvolution of the spectra into contributions from the individual repeats by tracing peaks through the spectra. Accordingly, a set of N- and C-terminally truncated fragments was designed by splitting $Y_IMR^1R^2R^3MA_{II}$ between position 41 and 42 of individual internal repeats. The series contained five N-terminally truncated fragments ($MR^1R^2R^3MA_{II}$, $R^1R^2R^3MA_{II}$, $R^2R^3MA_{II}$, R^3MA_{II}, MA_{II}) and three C-terminally truncated fragments (YMR^1, YMR^1R^2, $YMR^1R^2R^3$).

All fragments expressed with adequate yields, were soluble and could be purified to effective homogeneity. Representative [^{15}N, ^{1}H]-HSQC spectra of N-terminally truncated fragments are shown in Supp. Fig. S3 indicating that these fragments constitute well-folded proteins. Only spectra of $R^1R^2R^3MA_{II}$ displayed line-broadening, indicating the formation of oligomeric species. Parallel studies

revealed that the C-terminally truncated versions (e.g. $Y_IMR^1R^2R^3$, $Y_IMR^1R^2$ and Y_IMR^1) were generally unstable (data not shown), highlighting the importance of the A_{II}-cap for protein stability and correlating with the findings of Watson *et al*. [31].

[^{15}N,^{1}H]-HSQC spectra of the ^{15}N-labeled N-terminally truncated fragments showed considerable agreement with the spectrum of full-length $Y_IMR^1R^2R^3MA_{II}$. We hypothesized that signals from each repeat would successively appear close to, or directly at their final position when the length of the fragment was extended by each successive repeat. Signals were tracked from the shortest to the longest variant. Signals were assigned to repeats in the order of their appearance in spectra of fragments of increasing size. Figure 1 depicts the assignment strategy and shows an example of signals successively appearing in the Gly region of the spectra.

This strategy extended our assignment significantly, and stretches of sequence identified previously simply as part of "an" unrandomized helix H2 could now be mapped to a specific repeat. Most importantly we were able to distinguish between signals from the two identical helices H3 of the two unrandomized M repeats, and achieved assignments of at least parts of H3 in all repeats. We observed that, as expected, signals of residues close to the truncation site tend to move into their final position only after another repeat module had been added, in contrast to residues further away from the truncation site. We were able to effectively employ this strategy up to $R^2R^3MA_{II}$, after which the increased spectral complexity and broader lines of $R^1R^2R^3MA_{II}$ prevented reliable transfer of assignments. We therefore back-tracked assignments in the opposite direction from the full-length $Y_IMR^1R^2R^3MA_{II}$ to $MR^1R^2R^3MA_{II}$ in order to close the assignment gaps as far as possible. Using this method we were even able to assign the side chain indole protons of the three Trp residues present on the binding surface in repeats 1, 3 and 5. For $Y_IMR^1R^2R^3MA_{II}$ without NT, the backbone assignment coverage was extended from 20.6% to 36.8% and for the complex with NT from 32.4% to 44.9%. These initial assignments were sufficient to identify protein regions directly or indirectly affected by the binding of NT. The status of the assignments for backbone resonances in both the absence and presence of NT are summarized in Figure 2A and 2B, respectively.

Chemical shift perturbations (CSP) localize the peptide-binding site

CSPs clearly demonstrate that NT interacts with $Y_IMR^1R^2R^3MA_{II}$. Examples from spectral regions of Gly and Trp-indole resonances are depicted in Figure 3A and B.

A gradient in the magnitude of CSPs was observed across the protein with the largest changes occurring at the N-terminal end. As mentioned above, addition of NT also enabled the unambiguous assignment of residues 20 to 30 in the N-cap that were undetectable in the peptide-free sample due to peak broadening.

ArmRPs can be truncated and such fragments naturally reassemble to a full-length ArmRP[31]. Preliminary interaction studies reassembling various fragments of the $Y_IMR^1R^2R^3MA_{II}$ binder revealed that not all repeats in the original binder were involved in forming contacts to the peptide (data not shown). Briefly, N-terminal fragments containing the N-cap and different numbers of internal repeats, but no C-cap, were mixed with complementary C-terminal fragments containing only internal repeats and a C-cap (Fragment sequences see Figure S1) and binding of the peptide was tested by CSP. These studies confirmed that the N-terminal part of the protein contributes more towards peptide binding (Figure 2 and 3) and suggested that some internal repeats could be removed from the C-terminal end. This was verified by engineering constructs that systematically removed repeats, retaining both the N- and the C-cap. The resulting ArmRPs $Y_IMR^1R^2R^3A_{II}$, $Y_IMR^1R^2A_{II}$ and $Y_IMR^1A_{II}$ were all soluble, well-expressed, and yielded high quality [^{15}N, ^{1}H]-HSQC spectra (see Figures 3 D-F and S4).

CSP studies of these proteins revealed that the shortest construct, $Y_IMR^1A_{II}$, containing only two internal repeats, displayed only minor chemical shift changes indicating very weak binding. $Y_IMR^1R^2A_{II}$ was the smallest (22 kDa) construct displaying affinity for NT (see Figures 3E and S5, quantified by SPR Figure S6 and chemical shift mapping Table 1 and S7) that was comparable to $Y_IMR^1R^2R^3MA_{II}$ (32 kDa). Since the magnitude of CSPs remained relatively constant upon truncation and comparisons of K_d's determined from CSP titrations and SPR indicated no significant difference, further investigations were carried out with this reduced-size binder.

Assignments of $Y_IMR^1R^2A_{II}$ were achieved with much less ambiguity, largely due to the reduced size and the absence of a second identical M repeat, resulting in high quality spectra with significantly less peak overlap. In this case, almost all backbone resonances of the H3 helices could be assigned. Furthermore, some resonances from the N-cap were observed in the free form for the first time, and in the NT-complex

nearly the complete N-cap was visible. For $Y_IMR^1R^2A_{II}$ 82.8% of all non-Pro backbone resonances were assigned. In complex with NT, the coverage increased to 97.4% so that only five residues in loop regions and at the beginning of the N-cap remained unassigned (Figure 4). In addition, partial assignments for side chains of $Y_IMR^1R^2A_{II}$ in complex with NT for the putative binding interface were obtained from amide-anchored triple-resonance spectra such as (H)CC(CO)NH and H(CCCO)NH [32].

Interaction Surface of the Complex formed by $Y_IMR^1R^2A_{II}$ and NT mapped by CSP, PREs and Automatic Docking

Nearly complete backbone assignments of free and NT-complexed $Y_IMR^1R^2A_{II}$ now allowed the evaluation of the CSP data (Figure S5). The most strongly affected polypeptide backbone amide NH resonances $\left(\Delta\delta_{HN} > \overline{\Delta\delta_{HN}} + \sigma\right)$ arise for residues located at or interacting with the N-cap such as Gln19 ($\Delta\delta_{HN}$ ~ 0.60 ppm), Leu21 (~ 0.65 ppm) and Val51 (~ 0.33 ppm); as well as residues located in H3 of the internal repeats M (Ser80, ~ 0.33 ppm and Ser84, ~ 0.83 ppm), R^1 (Asn122, ~ 0.34 ppm) and R^2 (Ile166, ~ 0.39 ppm) (Figure S8A). In contrast, residues located in the C-cap are not perturbed. Some residues are only detectable in presence of NT, suggesting that protein regions that are in intermediate conformational exchange in ligand-free ArmRP become locked into one conformation upon binding NT. Such residues are localized in the N-cap (residues 16-18, 22-23, 32-42), in the vicinity of the hinge region connecting the N-cap to the first internal repeat (residues 47-49 and 57), and the conserved Asn ladder in repeats M and R^1 (residues 81 and 123), whereas the C-cap remains unaffected (Figure S8A). Thus, a consistent picture emerges from the CSP mapping indicating that NT interacts with $Y_IMR^1R^2A_{II}$ on the continuous interface spanned by H3 helices of the internal repeats, the strongest interactions occurring for repeats M and R^1; the C-cap remains unaffected (Figure 4).

While the CSP data strongly suggest a specific interaction involving the canonical binding surface of ArmRP, it fails to characterize the binding mechanism at the atomic level. To gain further insight, we docked a short NT fragment (NT7-13, with an acetylated N-terminal amino group) to a model of $Y_IMR^1R^2A_{II}$ via AutoDock Vina (see methods). Interestingly, the predicted ligand poses cluster into two opposite orientations that we term “parallel” and “antiparallel” based on the relative alignment

of the termini for ligand and receptor. The parallel pose correlates well with the CSP data and is energetically more favorable (Figure 4 and S9A), however an antiparallel conformation (see Supp. Mat. Figure S9B) could not be discarded at this stage.

Since no NOEs could be detected between the peptide and the protein, most likely due to an insufficiently long lifetime of the bound state, we decided to verify the binding location and determine the orientation of NT using paramagnetic relaxation enhancement (PRE) tags attached to NT. In this method, the attenuation of protein signals is solely related to the distance separating a given resonance from the unpaired electron in the spin-label and, in contrast to the CSPs, is not affected by induced conformational changes. NT was labeled with the nitroxyl PRE-tag *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate (MTSL) at peptide positions 1, 6 and 13 in order to bracket the four peptide residues $Pro7^{NT}$, $Arg8^{NT}$, $Arg9^{NT}$ and $Tyr11^{NT}$ previously identified as critical for binding [23], while minimizing direct interference with binding. Spin-labeling of NT at position 6 ("NT-K6C") resulted in distinct attenuations in the upper part of the binding interface, which surround the central repeat R_1 (Figure 5A), strongly supporting our hypothesis of a parallel binding mode. Indeed, predicted PREs (Figure 5B) based on an additional docking run of NT(7-13) labeled at position 6 with MTSL (*i.e*. NT-K6C-13) display significant correlation with the experimental PREs (Figure 5A).

A less strongly affected area around H3 of R^2 was identified by the spin label located at position 13 ("NT-L13C"), in line with the results from position 6 (see Supp. Mat. Figure 10B). Moreover, the non-localized distribution of mostly weak attenuations induced by a spin label at position 1 ("NT-Q1C") confirms that the N-terminal hexapeptide is not involved in a specific binding mechanism, consistent with previous binding data [23] (see Supp. Mat. Figure S10A).

Because the PRE effect is inherently long-range, we also attempted to identify the location of bound MTSL in more detail via *differential* chemical shift perturbation ($\Delta\Delta\delta$, see Methods) mapping of $Y_IMR^1R^2A_{II}$ complexed with either unlabeled NT or with NT coupled to quenched MTSL at position 6. Interestingly, the largest differential proton CSPs occur in the vicinity of Tyr116 (Figure S11 and S8B).

To conclude, in light of the NT conformations suggested by automatic docking, out of the two basic orientations only the *parallel* ligand binding mode is compatible with the experimental CSP and PRE data. We carried out MD simulations starting from

both docked orientations and found that for the parallel orientation the interactions of NT and $Y_IMR^1R^2A_{II}$ are in agreement with experimental data and that this binding pose is more stable (see Figure S12, S13 and 6). We therefore reason that the parallel orientation is more likely and present further analysis of NT in this orientation.

The NT Binding Mode at Atomic Resolution

Although the experimental data and docking results discussed above are consistent with each other and provide strong evidence that NT interacts with the canonical ArmRP binding surface in a parallel fashion, the approximations inherent in rigid-protein docking require further validation of the predicted binding modes. We therefore ran two independent explicit solvent MD simulations of the $Y_IMR^1R^2A_{II}$: NT(7-13) complex starting from the parallel orientation predicted by docking. The complex is stable over a 2-μs time scale but the NT(7-13) peptide shows remarkable flexibility in its C-terminal section. The side chains of $Pro7^{NT}$, $Arg8^{NT}$, $Arg9^{NT}$ and $Tyr11^{NT}$ form stable interactions with the protein, whereas $Pro10^{NT}$, $Ile12^{NT}$ and $Leu13^{NT}$ are involved in intra-peptide hydrophobic contacts and/or are partially exposed to the solvent (*cf*. movie). Both electrostatic and van der Waals interactions involving NT side chains contribute to binding (see Figure 6 and Table S4). The salt bridge between the guanidinium group of $Arg8^{NT}$ and the carboxyl of Glu158 is extremely stable (Figure 6B). In one simulation, the salt bridge between the side chains of $Arg9^{NT}$ and Asp43 exists – either directly or mediated by one or two water molecules – in about half of the trajectory (Figure 6C). In the second half of the trajectory, the side chain of $Arg9^{NT}$ moves away from Asp43 and interacts with Ser84 and Asn123. These interactions are consistently present in the second simulation as well (Figure 6C). The side chains of $Pro7^{NT}$ and Trp77 are optimally packed during the MD runs in a typical CH-π interaction [33], whereas the backbone O atom of $Pro7^{NT}$ is hydrogen-bonded to the Hγ1 atom of Ser80 (Figure 6A). The stacking of $Tyr11^{NT}$ and Phe126 is seen in both simulations, although in one of them it is transient and shows multiple events of formation and rupture (Figure 6D). In addition to the stacking interactions, there is a hydrogen bond between the Hη atom of $Tyr11^{NT}$ and the Oδ1 atom of Asn123, which is more persistent in one of the two simulations (Figure 6D). It seems that the peptide orientation and interactions in simulation 1 (solid lines in Figure 6) converge towards those observed almost from the beginning

of simulation 2 (dotted lines in Figure 6). This trend is consistent with the time evolution of the root-mean-square deviation (RMSD) of the Cα atoms of the NT(7-10) region of the peptide or the complete NT(7-13), calculated upon fitting MR^1R^2 repeats to the last structure from the second simulation (Figure S12). Residues 7-10 of NT(7-13) are well "anchored" and therefore have a lower RMSD than NT(7-13).

The simulations agree with previous experimental findings [23] and the PRE and CSP data. Concerning the latter data, the ring of $Pro7^{NT}$ packs against the aromatic ring of Trp77, while its Cα atom remains in close proximity to Ser80 (Table S4 and Figure 6) whose amide and Cβ resonances are both strongly perturbed ($\Delta\delta_{CB} \sim$ 6 ppm) in the presence of NT. The fact that no large amide CSP is observed for Glu158 ($\Delta\delta_{HN}$ = 0.041 ppm) may indicate that the side chain conformation is not significantly changed and also reflects the larger distance from the backbone amide to the actual point of interaction at the side chain head group. The conserved Asn-ladder does appear to be significantly involved in the interaction; $Arg8^{NT}$, $Arg9^{NT}$ and $Y11^{NT}$ interact with Asn123, whereas $Pro7^{NT}$ and $Arg9^{NT}$ interact weakly with Asn81 (Table S4). These Asn residues are strongly affected in the CSP data (Figure S8). Both $Arg8^{NT}$ and $Arg9^{NT}$ interact with Ser84, which is also strongly perturbed. Furthermore, the guanidium group of $Arg9^{NT}$ is transiently involved in a salt bridge with Asp43 (moderate CSP).

We noticed a network of aromatic residues in the upper part of the NT binding interface involving Trp77, Tyr116, and Trp161. It is likely that the π–π interactions of this network contribute to the structural stability of the protein. On the other hand, they may be involved in formation of weak CH–cation interactions with NT residues and provide a hydrophobic surface against which $Pro7^{NT}$ can pack.

Interestingly, Ile166 is perturbed (Figure S8) despite its remote location from the binding site. Based on MD simulations, Phe126 forms a transient π-stacking with Trp161 (Figure S14B), which, in the absence of NT, brings the Phe aromatic ring close to Ile166 (Figure S14A). Furthermore, in the presence of NT both Phe126 and Trp161 interact with $Tyr11^{NT}$ (Figure 6D). We hypothesize that as a result, Ile166 is no longer sufficiently close to experience the ring-current shift of Phe126.

Mutations in the N-cap modulate NT binding and affect packing of the N-cap against the first repeat

The exact nature of the N-cap was found to have a dramatic effect on the capability of the various proteins to bind NT; the stabilized mutants of the original binder VG_328 no longer binding NT [23]. Therefore, the effect on peptide binding of the differences between the Y_I- and the Y_{II}-cap was investigated in detail using a series of mutants. The complete change from a Y_I-cap to a Y_{II}-cap includes three mutations: V34R, R37S and R42Δ. All possible single and double mutations, as well as the triple mutation (= Y_{II}-cap), were introduced into $Y_IMR^1R^2A_{II}$. Moreover, R42A was introduced as single point mutation to distinguish effects due to shortening the loop between the N-cap and the first internal repeat (R42Δ), from those caused by removing the Arg42 side chain as a potential point of interaction with NT. Additionally, an E46A mutant was probed for the effect of removing a negatively charged residue from the N-cap/M-repeat loop, as NT contains several positively charged residues.

The original full-length binder $Y_IMR^1R^2R^3MA_I$ (VG_328) and its binding-competent version with stabilized C-cap $Y_IMR^1R^2R^3MA_{II}$ were used as reference proteins. CSPs from the titration of ^{15}N-labeled protein with NT were used to determine K_d values of all protein variants as described in the Materials and Methods. In addition, we titrated $Y_IMR^1R^2A_{II}$ with NT(7-13), a truncated version of NT. All results are summarized in Table 1. Exemplary fitted raw data for $Y_IMR^1R^2A_{II}$ are shown in Supp. Figure S7.

Interestingly, the K_d of $Y_IMR^1R^2A_{II}$ with NT(7-13) was 12 ± 5 μM and hence in the same range as the full-length peptide, indicating that the first 6 residues of NT do not contribute to binding, as suggested previously [23].

The V34R and R42A point mutations were found to have the smallest impact on binding, with the K_d remaining in the same range as observed for $Y_IMR^1R^2A_{II}$. The R42Δ mutation increased the K_d significantly (5 fold), while R37S was found to be the most disruptive point mutation increasing the K_d by a factor of 12.5. Combinations of the single mutations showed synergistic effects, e.g. V34R/R37S with a factor of 15 and V34R/R42Δ with a factor 6. The effect of all three mutations present in $Y_{II}MR^1R^2A_{II}$ led to an increase in K_d by a factor of 20.5. The K_d's of the original binder VG_328 ($Y_IMR^1R^2R^3MA_I$), the binder with the stabilized C-cap,

$Y_IMR^1R^2R^3MA_{II}$, and the optimized minimal binder $Y_IMR^1R^2A_{II}$ were all very similar.

Additionally, all $Y_IMR^1R^2A_{II}$ mutants and VG_328-based reference proteins described above were assessed for interaction with NT by ELISA (see Figure S17). For $Y_IMR^1R^2R^3MA_{II}$ and $Y_IMR^1R^2A_{II}$, binding constants were also confirmed by surface plasmon resonance (SPR) studies (for results see Table 1). Biotinylated NT was immobilized on the chip as described in Materials and Methods. ELISA results of the $Y_IMR^1R^2A_{II}$ variants and the VG_328 derived reference proteins were in good agreement with CSP-based K_d results, corroborating the trends described above.

The affinity of $Y_IMR^1R^2A_{II}$ and $Y_IMR^1R^2R^3MA_{II}$ for NT was determined by SPR at 8 °C (*cf*. Materials and Methods) to be 14 μM and 18 μM, respectively. Earlier studies determined a K_d of 7 μM for the original binder VG_328 ($Y_IMR^1R^2R^3MA_I$) with the A_I-type C-cap at 4 °C and a similar experimental set-up [23]. It should be noted that for binders in the μM range, K_d values from SPR have higher errors compared to those obtained from NMR experiments.

Protein variants with sequences of internal repeats identical to $Y_IMR^1R^2R^3MA_{II}$ but containing stabilized versions of the N-cap are not capable of binding the NT peptide (*vide supra*). NMR data indicate that the N-cap is not well folded, and the absence of signals is indicative of molten-globule type behavior. Since the previous MD simulations of proteins with this N-cap showed that the latter does not pack well against the remainder of the protein [25], we performed a 2-μs MD simulation of $Y_IMR^1R^2A_{II}$ in the absence of NT. The simulation utilized the above-described structural model as the starting conformation. We observed that the N-cap and the loop containing residues 38-48, which connects the N-cap and the first internal repeat, had considerable flexibility. Moreover, the relative orientation of helix 2 of the N-cap and helix 3 of the first internal repeat, as characterized by the angle θ, was monitored along the trajectories of the apo proteins $Y_IMR^1R^2A_{II}$ and $Y_{III}M_3A_{II}$ (see Figure 7). The N-cap of $Y_IMR^1R^2A_{II}$ shows a shift in the distribution of the theta angle towards larger values with respect to $Y_{III}M_3A_{II}$. Through the larger rotation of the entire N-cap the loop connecting the N-cap with the first repeat is shifted into closer proximity to the binding surface, possibly accounting for the PRE attenuations around residue Gly44. The angle θ is also larger in $Y_IMR^1R^2A_{II}$ than $Y_{III}M_3A_{II}$ in control simulations performed using a different force field and at 330K for enhanced sampling (Figure

S15).

We further investigated the effects of different N-cap mutations on protein stability in the absence of NT (see Figure 8A). Root mean square fluctuations (RMSF) of C_α atoms obtained from MD simulations reveal that the effects of the N-cap mutations are restricted to the 50 N-terminal residues of the protein. $Y_IMR^1R^2A_{II}$, $Y_IMR^1R^2A_{II}$_V34R and $Y_IMR^1R^2A_{II}$_R37S display a similar pattern of fluctuations, with the highest fluctuations in the loop region between the N-cap and the first repeat (residues 41-46) (see Figure 8A). In this region, fluctuations are slightly higher in $Y_IMR^1R^2A_{II}$ than in the mutants. The R42Δ deletion has a similar effect as the V34R and R37S mutations, only fluctuations in the loop region (residues 41-46) are somewhat lower (data not shown). Importantly, the RMSF profiles indicate that both $Y_IMR^1R^2A_{II}$ and $Y_IMR^1R^2A_{II}$_R37S are stabilized by the presence of NT, except for the loop between H1 and H2 helices of the N-cap of the $Y_IMR^1R^2A_{II}$_R37S:NT(7-13) complex (Figure 8B).

The second half of the H2 helix of the N-cap, the loop connecting the N-cap and the first repeat, and the H3 helix of the first repeat appear to be most stabilized by NT.

We have also observed that the distance between the guanidinium group of Arg37 and the aromatic ring of Trp77 is mostly more than 2 Å larger in the apo state of the protein (see Supp. Mat. Figure S16), indicating that Arg37 may be of importance in organizing the aromatic network in the complex. Moreover, in all MD runs an interaction between Met17 and Phe36 is present most of the time, possibly contributing to the stability of the N-cap (Figure 9). Valley et al. [34] observed such stabilizing Met-aromatic motifs in approximately one-third of all known protein structures. However, in $Y_IMR^1R^2A_{II}$_R37S with NT(7-13) bound, the distance between Met17 and Phe36 increases from ~5 Å to ~7 Å in the second half of the 2 μs MD simulation. This may be a reason for the larger RMSF in the loop region between the helices H1 and H2 of the N-cap (Figure 8B). In the first half of the 2-μs MD simulation, the RMSF are similar in magnitude to the other RMSF shown in Figure 8B.

To probe for possible differences in conformational stability between the free and NT-bound states we additionally measured the exchange rate of $Y_IMR^1R^2A_{II}$ amide protons in two independent series of MEXICO [35] experiments. As expected, the presence of NT leads to a generally lower exchange rate. Mapping the effect on

the backbone of $Y_IMR^1R^2A_{II}$ uncovers that some of the most affected residues (Gln19, Gln20, Leu21, Gln29, Leu30, Ser66, Asn68) cluster in the vicinity of the N-cap hinge region (Figure S18). This is in line with our finding that addition of NT results in a stabilizing effect for the N-cap via the network of cation-π, π-π and CH-π stacking interactions as discussed earlier.

Discussion

The present study highlights some unique challenges that must be overcome when analyzing weak protein-peptide interactions. While the individually obtained experimental data are insufficient to allow unambiguous interpretation of the binding mode, the overall picture derived from combining various approaches provides much insight to help drive forward the design and construction of engineered ArmRPs (see Figure 10).

A problem at the onset of the project was the limited stability of the protein versions available at that time, which, however, was rapidly improved through introduction of stabilized C-caps (Q292L and F293Q = A_{II}) [25]. However, despite favorable biochemical properties the highly repetitive sequence did not allow extensive assignments in $Y_IMR^1R^2R^3MA_{II}$.

A strategy pursued in this study to simplify the assignment problem was to use fragments. It was noted elsewhere [31] that N-terminally truncated armadillo repeat proteins are stable and resulted in [^{15}N,^{1}H]-HSQC spectra that were largely superimposable with spectra of their full length parent. In contrast, C-terminally truncated fragments displayed molten globule-like behavior, regardless of their length. This suggests that the A_{II} C-cap plays a crucial role for protein stability, an effect that is transferred through the complete protein. Conversely, the Y_I N-cap appears to be far less stable, a view that is supported by our observation that many signals from the N-cap were missing in the [^{15}N,^{1}H]-HSQC spectra. While subsequent engineering of the N-cap has largely solved this problem [25], the original selections that led to the NT-binder VG_328 had been performed with a library containing the Y_I-cap [23], and indeed, this particular cap was required to maintain binding.

A similarly decisive role of cap stability for overall protein stability was noted earlier in the class of Ankyrin repeat proteins [28]. The systematic truncation of N-terminal repeats from $Y_IMR^1R^2R^3MA_{II}$ by one repeat at a time allowed us to increase the backbone resonance assignments to an extent that revealed that NT binding to $Y_IMR^1R^2R^3MA_{II}$ had no effect on the resonances of the C-cap and the last internal repeat.

Once the approximate binding location of NT was known, it became possible to eliminate whole repeats unimportant for peptide binding. The resulting optimized minimal binder $Y_IMR^1R^2A_{II}$ retained full NT binding properties and displayed

improved spectra. Since $Y_IMR^1R^2A_{II}$ represents a much smaller target – 22 kDa instead of 32 kDa – and because assignments were easier in the absence of two identical unrandomized M-repeats (which had been added to the library to improve protein stability [23]), we were able to achieve near-complete and unambiguous backbone assignment of the whole protein, as well as partial side-chain assignments for the binding interface formed by helix 3 of each repeat. Thereby, detailed analysis of the interaction with NT by CSP and PREs became possible. The pruning of unnecessary repeat modules from established binders at no expense in binding affinity represents a novel strategy in the development process of repeat protein engineering.

CSP experiments of $Y_IMR^1R^2A_{II}$ with NT revealed substantial changes in the binding interface and the N-cap, and smaller changes in the hinge regions between helices. The effects were spread over a much larger area than expected. We interpreted this as a combination of direct effects due to peptide binding and indirect effects due to structural rearrangements involving the N-cap. We hypothesize that the N-cap is locked into one position only upon binding of NT, causing a series of strong CSPs in the interface between the N-cap and the first internal repeat. These stabilizing effects are propagated from the cap through the whole protein leading to minor CSPs in hinge regions.

The undiminished interaction of NT7-13 with $Y_IMR^1R^2A_{II}$ in NMR studies indicated that the N-terminal hexapeptide of NT does not contribute significantly to binding, consistent with previous experiments [23]. PRE studies also showed that the central region of the peptide is more rigidly located in the complex, whereas the N- and C-termini sample a number of conformations in non-contiguous regions of the protein surface. Since the location of the nitroxyl moiety relative to the backbone is intrinsically less well-defined, and because intermolecular NOEs could not be detected due to the relatively low binding affinity, the NMR data alone do not allow NT to be placed on the $Y_IMR^1R^2A_{II}$ binding surface unambiguously. In order to progress with defining consistent poses and conformations, we therefore turned to computational methods such as docking and MD simulations.

Automatic docking suggested two poses for NT binding to $Y_IMR^1R^2A_{II}$, a parallel and an antiparallel binding mode. The antiparallel pose is clearly not in agreement with the majority of NMR data, and can therefore only be populated to a small extent, if at all. In contrast, the parallel binding mode results in a large number of favorable interactions between peptide and protein, and is in agreement with CSPs and PREs. In

the MD simulation starting from the parallel pose, which in turn was proposed by docking, the salt bridge involving NT residues Arg8^{NT} with Glu158 is stable while the one between Arg9^{NT} with Asp43 is only transiently formed. Moreover, the side chains of Pro7^{NT} and Trp77 are packed favorably. These interactions are also compatible with the results from the Ala-scan performed earlier on NT [23]. Interestingly, the suggested binding mode does not mirror the canonical binding mode observed in naturally occurring Armadillo repeat proteins [36].

Nonetheless, it appears that NT does utilize the conserved asparagine ladder to some extent, albeit *via* side chain contacts. The central part of the peptide is bound more tightly and the N-terminus makes only transient interactions with other parts of the $Y_IMR^1R^2A_{II}$ binding interface. The binding hypothesis from this work is supported by the results from the Ala-scan of NT. It is also in agreement with the fact that the binder was developed by pre-panning against the first 5 residues of NT during ribosome display selections [23].

In summary, we provide evidence that the central part of NT (residues 7 to 11) makes contacts with the binding interface presented by helices 3 of $Y_IMR^1R^2A_{II}$ as intended in the original design. We successfully reduced the size of the original binder from 32 to 22 kDa without loss of binding competency, and confirmed that Pro7^{NT}, Arg8^{NT}, Arg9^{NT} and Tyr11^{NT} are key peptide residues for binding.

An unexpected observation in this study was that the protein mutants incorporating the stabilized N-caps no longer bound NT. Initially it was unclear whether this effect was due to removal of residues that form contacts with NT, or whether the geometry of the binding interface was altered and incompatible with NT binding. Interestingly, the three N-cap mutations, V34R, R37S and R42Δ, that reduce the binding affinity for NT (albeit to variable extent), all affect positioning of arginine residues. We studied the mutations individually to determine which of them were responsible for increased protein stability and which were crucial for NT binding. Titrations of NT against $Y_IMR^1R^2A_{II}$ and its N-cap variants enabled us to identify R37S as the most disruptive single mutation for peptide binding, increasing the K_d from ~18 μM to about 224 μM (factor 12.5), whereas R42Δ increased the K_d by a factor of 5 (see Table 1). Combinations of the single mutations showed additive effects, with the triple mutant (V34R, R37S, R42Δ) in $Y_{II}MR^1R^2A_{II}$ displaying a 20.5-fold increased K_d.

Surprisingly, no direct contacts are formed between Arg37 and NT7-13 in our model of the complex. Instead, Arg37 is revealed as an important pivot for organizing the

network of aromatic residues surrounding the binding site. Indeed, its guanidinium moiety pre-orientates the Trp77 indole ring via cation-π stacking and thus facilitates the packing of $Pro7^{NT}$ against the side chain of the latter.

Overall, our data provide a plausible explanation for the binding in that its mechanism appears to rely on two specific aspects, namely (i) packing of the N-cap against the first internal repeat, which to some extent (ii) pre-organizes the aromatic network surrounding the interaction site.

Conclusions

Undoubtedly, crystal structures of larger proteins provide fast access to structural information of protein-peptide complexes, and this is particularly true for repeat proteins. However, in early stages of such projects, when a new class of proteins is being developed, binding affinities are low and protein binders may still contain flexible parts hampering crystallization. Often, these binders remain poorly characterized due to the lack of proper methodology to investigate the details of these low-affinity complexes. Herein we have developed a highly interdisciplinary approach combining mutagenesis, heteronuclear NMR spectroscopy and atomistic simulation methods supported by other biophysical tools. We believe that this approach can be a powerful strategy for analyzing difficult targets such as low affinity binders with multiple binding modes. We have also demonstrated that, even based on information from limited NMR assignments, protein sequences can be modified to yield proteins with superior characteristics that may eventually be amenable to high-resolution structural studies. Most importantly, this limited information is sufficient to drive the project forward and to verify original hypotheses about the binding mode of the ligand in designed binders. Particularly in the early stages of this type of project, it is of the utmost importance to ensure that it is on the correct track.

Materials & Methods

Nomenclature

The ArmRPs in this study contain consensus repeats (M) and randomized internal repeats (R) based on the previously described $\overline{M}$-type (for more details see [23, 25]).

In the case of the selected binder VG_328, the protein contains 3 randomized library modules termed R^1, R^2 and R^3. The N-terminal capping repeat, derived from yeast importin-α is termed "Y". The C-terminal capping repeat, was artificially designed [25] termed "A". The number of identical repeats in a protein is indicated as a subscript, e.g. a protein with five identical internal consensus repeats is called YM_5A. To distinguish different design versions of capping repeats, the caps are labeled with additional subscripts in roman numerals, e.g. $Y_{II}M_5A_{II}$. In the presented nomenclature, the used binder VG_328 is denoted as $Y_IMR^1R^2R^3MA_I$. All cap and internal repeat sequences used in this study are shown in Figure S1.

Cloning

Experiments were performed according to standard procedures [37] unless stated otherwise. Oligonucleotides were purchased from Microsynth AG (Balgach, Switzerland); for a complete list of all used oligonucleotides see Table S2. Enzymes and buffers were from New England Biolabs or Fermentas (Lithuania). *E. coli* strain XL1-blue (Genotype: *recA1, endA1, gyrA96, thi-1, hsdR17*(r_K^-, m_K^+), *supE44, relA1, lac*, [*F', proAB, lacIqZΔM15::Tn10(tet^r)*], Stratagene, California, USA) was used for cloning. Further details on the cloning procedure are provided in the Supp. Mat.

Expression of Proteins

Proteins were expressed in *E. coli* M15 [pREP4] in LB medium for unlabeled protein and in M9 minimal medium supplemented according to the desired isotopic labeling with ^{15}N-NH_4Cl, ^{13}C-glucose and ^2H_2O as described previously [28,31].

For the expression of ^{15}N, ^{13}C, ^{2}H-labeled proteins 5 mL LB_{D2O} overnight starter cultures were used to inoculate 50 mL D_2O minimal medium pre-cultures, which were incubated overnight to increase cell density before being used to inoculate the final culture at a volumetric ratio of 1:20. Expression was induced at $OD_{600} = 0.6$ and carried out for 16 h at 37 °C. Using ^{2}H, ^{13}C-glucose the final level of deuteration was about 90%.

Protein Purification and Characterization

Cell pellets were resuspended in TBS_{500} [50 mM Tris-HCl, 500 mM NaCl, 5% (v/v) glycerol, pH 8.0] and purified as previously described [28, [31].

For complexation with NT, the His_6-tag of C-terminal ArmRP fragments was removed by rTEV protease at a molar ratio of 1:30 as previously described [31].

After IMAC purification ArmRPs and fragments were further purified by preparative size exclusion chromatography (SEC) in PBS_{150} pH 7.4 with 2% (v/v) glycerol on a S75 16/60 HiLoad column (GE Healthcare). Protein size and purity were checked by 15% SDS-PAGE. Proteins were further analyzed by ESI mass spectrometry to verify the exact mass and determine the degree of isotopic labeling.

Analytical SEC was carried out on a Superdex 200 5/150 GL (Pharmacia) column on an ÄKTA HPLC system in PBS_{150} (50 mM phosphate and 150 mM NaCl, pH 7.4, 2% glycerol) ArmRPs have been shown to elute at a higher apparent size than the calculated monomeric weight suggests. This is a result of their elongated shape and greater effective hydrodynamic ratio [36].

NMR Spectroscopy and Data Evaluation

Spectra were recorded in PBS_{150} buffer (150 mM NaCl, 50 mM Na-phosphate, 2% (v/v) glycerol, pH 7.4) supplemented with 10% D_2O, 1 mM TMSP-d4, 0.0 1% NaN_3 and 2% (v/v) glycerol. Protein solutions were concentrated to 0.2-1.0 mM for NMR measurements. NMR data were recorded at 32°C on Bruker AV-600 or AV-700 MHz spectrometers equipped with triple-resonance cryoprobes. Data were processed in TOPSPIN 2.1 and analyzed with CARA [38]. Resonances were calibrated relative to the proton water resonance at 4.63 ppm, the ^{15}N and ^{13}C scales were calculated indirectly (conversion factors $^{15}N = 0.10132900$, $^{13}C = 0.25144954$). Experiments were selected from the Bruker standard pulse sequence library, and used pulsed-field gradients, sensitivity-enhancement schemes, and water suppression through coherence selection [29,30].

For backbone assignments, ^{15}N, ^{13}C, ^{2}H-labeled proteins were used. Deuterium decoupling was applied during relevant ^{15}N- or ^{13}C-evolution periods or delays. Sequential amide spin systems were linked via matching carbonyl (HNCO/HN(CA)CO experiments) and Cα and Cβ resonances (HNCACB/HN(CO)CACB experiments). Additionally, HN(CACO)NH and ^{15}N-3D-NOESY experiments provided sequential correlations of nitrogens and protons of amide groups, respectively [28]. Initially, sequential assignments were made automatically using the program MARS [39], then manually checked and completed. For side chain assignments, constant-time [^{13}C,^{1}H]-HSQC experiments

combined with (H)CCH-TOCSY and ^{13}C-resolved aliphatic or aromatic-NOESY experiments of uniformly ^{15}N, ^{13}C-labeled protein were used. All assigned chemical shifts of $Y_IMR^1R^2A_{II}$ in the presence of NT have been deposited in the BMRB database under accession code 25367.

Chemical Shift Perturbation (CSP) Experiments

Chemical shift mapping was used to probe for conformational changes in the protein upon peptide binding and to investigate direct protein-peptide interactions. Shift deviations ($\Delta\delta$) for $Y_IMR_1R_2R_3MA_{II}$ and $Y_IMR_1R_2A_{II}$ upon complex formation were taken from [^{15}N, ^{1}H]-HSQC spectra recorded in absence and presence of 2 molar equivalents of NT and quantified using the formula

$$\Delta\delta_{obs} = \sqrt{\left(\Delta\delta_{^1\mathrm{H}}\right)^2 + \left(\left|\frac{\gamma_N}{\gamma_H}\right| \cdot \Delta\delta_{^{15}\mathrm{N}}\right)^2}$$

where $\Delta\delta_{1H}$ and $\Delta\delta_{15N}$ correspond to the backbone amide chemical shift differences; γ_H and γ_N correspond to the gyromagnetic ratios for the proton and nitrogen resonances, respectively [40]. Additionally, we define a *differential* chemical shift perturbation ($\Delta\Delta\delta \equiv \Delta\delta_{cond2} - \Delta\delta_{cond1}$) to directly compare perturbations that were observed under two separate conditions. Note that $\Delta\Delta\delta$ refers to a *net* difference between quadratically normalized CSP for two conditions and can thus take positive as well as negative values. Thereby, we estimate the impact of adding a spin label (*vide infra*) to the NT peptide by quantifying the *differential* CSP of $Y_IMR_1R_2A_{II}$ complexed with either unlabeled NT ($\Delta\delta_{cond1}$), or with NT coupled to quenched MTSL ($\Delta\delta_{cond2}$).

Determination of Dissociation Constants (K_d) by NMR using CSP

NT binding to $Y_IMR_1R_2A_{II}$ was detected from perturbations of [^{15}N,^{1}H]-HSQC spectra by monitoring the chemical shift changes of protein backbone amides as a function of ligand concentration. A total of 5 equivalents NT or NT7-13 peptide solution were successively added to 250 µM protein samples in PBS_{150} buffer (150 mM NaCl, 50 mM Na-phosphate, pH 7.4) supplemented with 2% (v/v) glycerol, 1 mM TMSP-d_4 and 0.01% NaN_3. Assuming single-site binding for a system in fast exchange, quadratically weighted amplitudes of ^{1}H and ^{15}N chemical shift differences

at each titration step *i* were combined and fitted by non-linear regression analysis as [40]:

$$\Delta_{cal}^{i} = \Delta_{\infty} \frac{([P]_{total} + [L]^{i} + K_D) - \sqrt{([P]_{total} + [L]^{i} + K_D)^2 - 4([P]_{total} \cdot [L]^{i}}}{2 \cdot [P]_{total}}$$

using an algorithm implemented in MatLab as previously described [41]. Almost complete ligand saturation, ranging from 70% (K_d > 200 μM) to >95% (K_d < 30μM), was consistently achieved in the last ligand addition step of each titration series. Multiple binding curves derived for individual resonances were averaged to yield more precise K_d values for the interaction of each construct with NT.

Paramagnetic Relaxation Enhancement (PRE) Experiments

Cysteine mutants of the NT peptide (NT-Q1C, -K6C and -L13C) were obtained from Anaspec (Fremont , CA, USA), dissolved in PBS_{150} (50 mM Na-phosphate, 150 mM NaCl, pH 7.4) and incubated with a 2× molar excess of TCEP for 30 min at room temperature. A 10× molar excess of the PRE-tag MTSL (CAS: 81213-52-7,TRC, Toronto) dissolved in DMSO was added, and the pH adjusted to 9 using 1 M NaOH. The reaction mix was incubated for 2 h in the dark at room temperature with vigorous shaking. Complete labeling was confirmed by ESI mass spectrometry. Labeled peptides were purified in H_2O by SEC using a 30/10 peptide column (GE Healthcare), lyophilized, dissolved in PBS_{150} (50 mM Na-phosphate, 150 mM NaCl, pH 7.4) and added at 2× molar excess to NMR samples containing uniformly ^{15}N, ^{13}C-labeled $Y_IMR^1R^2A_{II}$. Two sets of [^{15}N,^{1}H]-HSQC (water flip-back) and [^{13}C,^{1}H]-HSQC (aliphatic and aromatic) experiments were recorded using relaxation delays of 2 s, in which the diamagnetic reference was obtained by addition of 10 equivalents of ascorbic acid. The inactivated sample was incubated at room temperature for 1 h and the pH readjusted to pH 7.4 using 1 M NaOH before recording the reference spectrum. The ratio of the signal intensity $MTSL_{active}$: $MTSL_{inactive}$ was used as an indicator of spatial proximity of the PRE-tagged peptide side chain to the attenuated residues of the protein.

Measurement of Differential Amide Proton Exchange Rates

Amide proton exchange rates in the presence and absence of NT were derived from a series of MEXICO [35] experiments using doubly matched ^{13}C and ^{15}N filters. Data measurement and evaluation followed procedures previously published by us[28].

Docking of NT to $Y_IMR^1R^2A_{II}$

Models of the NT7-13 peptide fragment were docked to $Y_IMR_1R_2A_{II}$ via AutoDock Vina 1.1.2 [42] in a free and unrestricted fashion, which uses a highly optimized algorithm to efficiently predict flexible ligand conformations on macromolecular receptor targets. A molecular model of the NT peptide fragment P7-R8-R9-P10-Y11-I12-L13 ("NT7-13", N-term acetylated) was constructed in PyMOL [43]. In order to analyze the effect of attaching a spin label to NT residue K6, the model was modified in Chem3D (Cambridgesoft) to incorporate an MTSL entity at its N-terminus, yielding "NTK6C-13". Both models were energy minimized using the MM2 force field [44] and prepared for docking with AutoDockTools 1.5.6 [45]. A torsion tree encompassing 31 rotatable bonds (maximum number allowed by the AutoDock algorithm) was defined for both ligands. To comply with this limit, all bonds of the C-terminal residue L13 as well as the bonds in the guanidium groups of residues R8/R9 had to be rigidified for ligand NTK6C-13. Considering that in the MD simulations the sidechain of Leu-13 moves freely as it points away from the binding interface we felt this was justified. Relaxed PDB coordinates were extracted from the MD trajectory of the modeled $Y_IMR_1R_2A_{II}$ protein in explicit water (*cf*. section on MD simulations). These were then regularized with MolProbity [46] and WHAT IF [47] followed by merging of nonpolar hydrogens and addition of Gasteiger atomic charges calculated using the PARSE force field [48] at pH 7.4. The docking space was defined as a grid of 32 × 28 × 30 Å, centered at repeat R_1 to encompass the entire binding surface spanned by the repeating H3 helices (Figure S1). Within that space, blind docking was carried out with AutoDock Vina using an exhaustiveness value of 128; 20 poses were calculated for NT7-13 and two additional poses for NT6C-13. The lowest energy docked conformers served as starting coordinates for the MD simulations.

Molecular Dynamics Simulations

All MD simulations were carried out in explicit water at constant temperature (310 K) and constant pressure (1 bar) using a velocity-rescaling thermostat and Berendsen pressure coupling [49, 50]. Periodic boundary conditions were applied in all three dimensions. Coulomb and van der Waals interactions were cut off at 1 nm. The long-range electrostatic interactions were treated by the Particle Mesh Ewald method [51]. The simulations of the $Y_IMR^1R^2A_{II}$-NT(7-13) complex were started from the parallel or antiparallel orientation as predicted by docking (Table S3). The N-terminus of NT was acetylated, whereas the C-terminus was negatively charged. The protonation state of the side chains was chosen to reflect the experimental pH 7.4: aspartate and glutamate side chains and the C-terminal carboxyl group were negatively charged, lysine and arginine side chains and the N-terminal amino group were positively charged, and histidine residues were kept neutral. Each system was solvated in a dodecahedral box of TIP3P water molecules, with the box edge at a distance of at least 1.2 nm from the protein surface. Ions (Na^+ and Cl^-) were added to neutralize the total charge of the system at the concentration of 150 mM. The energy of the system was minimized, using a steepest descent algorithm, before the system was equilibrated in a 0.1 ns position-restrained simulation at constant molecular number, volume, and temperature (NVT). Then, a 0.9 ns position-restrained simulation at constant molecular number, pressure (1 bar) and temperature (NPT), with positional restraints on protein and peptide when present, was carried out to equilibrate the pressure. For the protein-peptide systems, an additional 50-ns simulation with distance restraints on the peptide backbone was performed before the onset of 2 μs unrestrained NPT simulations. The simulations were carried out using the GROMACS software version 4.5.5 with the CHARMM36 force field [52] and the TIP3P potential for water molecules [53].

In all simulations, the LINCS algorithm was used to fix the length of all bonds [54]. Virtual sites were used for removing fastest degrees of freedom, which allowed an integration time step of 5 fs.

All structural models shown in this work were established based on PDB coordinates of experimental crystal structures of natural and designed ArmRPs by sequence adaptation, repeat merging and relaxation in Rosetta [55]. The model for $Y_IMR^1R^2R^3MA_{II}$ is based on the natural yeast karyopherin-α structure (PDB ID:

1EE4 [56]), the model for $Y_IMR^1R^2A_{II}$ is based on the designed consensus ArmRP $Y_{III}M_3A_{II}$ (PDB ID: 4DB6 [57]).

Determination of Dissociation Constants (K_d) by Surface Plasmon Resonance (SPR)

SPR experiments were carried out on a BIACORE 3000 (GE Healthcare Biosciences, Pennsylvania, USA) with PBS-T [50 mM Na-phosphate, 150 mM NaCl, 0.01% Tween-20, pH 7.4] as running buffer. 10 RU (response units) of synthetic, biotinylated NT were immobilized on a streptavidin-coated SA-chip (GE Healthcare Biosciences). Interactions of NT with $Y_IMR_1R_2R_3MA_{II}$ and $Y_IMR_1R_2A_{II}$ were measured at increasing concentrations of protein (0.06-200 µM, flow rate 50 µL / min, 50 µL injections, 5 min dissociation buffer flow). Measured values were corrected by subtraction of a reference signal from an uncoated cell. Due to fast equilibration of the system, plateau values were used to determine the dissociation constant (Scrubber, BioLogic software).

ELISA

MaxiSorp 96-well plates (Nunc) were coated with NeutrAvidin (100 µL per well, 66 nM, overnight, 4 °C). Wells were blocked with 300 µL of 1×PBS-TB (50 mM phosphate and 150 mM NaCl, pH 7.4, 0.3% BSA, 0.1% Tween-20) 1 h at room temperature. Biotinylated target peptide ([Biotin]-[6-amino-caproic acid]-[β-Ala]$_2$-NT) was immobilized (100 µL per well, 200 nM, 1 h, 4 °C) in PBS-TB. Proteins were dissolved in PBS-B (50 mM phosphate and 150 mM NaCl, pH 7.4, 0.3% BSA), all washing steps were carried out in PBS-TB. Plates were incubated with target protein (100 µL per well, 200 nM, 1 h, 4 °C). Wells were washed three times with 300 µL of 1× PBS-TB and incubated with anti-RGSH$_6$ mouse antibody (1:5000 in 1× PBS-TB, 1 h, 4 °C; Qiagen, Germany) as primary antibody. Plates were washed as described above and incubated with a goat anti−mouse IgG alkaline phosphatase conjugate (1:10,000 in 1× PBS-BT, 1 h at 4 °C, Sigma) as secondary antibody. Signals were developed with disodium 4-nitrophenyl phosphate (100 µL per well, 3 mM, 2 h, 37 °C, Fluka, in 50 mM $NaHCO_3$, 50 mM $MgCl_2$). Absorbance at 405 nm was measured with a Perkin Elmer HTS 7000 Plus plate reader (Reference absorbance wavelength 540 nm was deducted).

Acknowledgments: We acknowledge financial support by the SINERGIA program of the Swiss National Science Foundation (grant no. 122686).

References:

1. Hoogenboom, HR. Selecting and screening recombinant antibody libraries. Nature Biotech 2005;23:1105-1116.
2. Binz, HK, Amstutz, P, Plückthun, A. Engineering novel binding proteins from nonimmunoglobulin domains. Nat Biotechnol 2005;23:1257-1268.
3. Boersma, YL, Plückthun, A. DARPins and other repeat protein scaffolds: advances in engineering and applications. Curr Opin Biotechnol 2011; 22:849-857.
4. Caravella, J, Lugovskoy, A. Design of next-generation protein therapeutics. Curr Opin Chem Biol 2010;14:520-528.
5. Hosse, RJ, Rothe, A, Power, BE. A new generation of protein display scaffolds for molecular recognition. Protein Sci 2006;15:14-27.
6. Löfblom, J, Frejd, FY, Stahl, S. Non-immunoglobulin based protein scaffolds. Curr Opin Biotechnol 2011;22:843-848.
7. Almagro, JC. Identification of differences in the specificity-determining residues of antibodies that recognize antigens of different size: implications for the rational design of antibody repertoires. J Mol Recognit 2004;17:132-143.
8. Kuriyan, J, Cowburn, D. Modular peptide recognition domains in eukaryotic signaling. Annu Rev Biophys Biomol Struct 1997;26:259-288.
9. Coates, JC. Armadillo repeat proteins: beyond the animal kingdom. Trends Cell Biol 2003;13:463-471.
10. Blatch, GL, Lassle, M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. BioEssays 1999;21:93-939.
11. Smith, TF, Gaitatzes, C, Saxena, K, Neer, EJ. The WD repeat: a common architecture for diverse functions. Trends Biochem Sci 1999;24:181-185.
12. Andrade, MA, Bork, P. HEAT repeats in the Huntington's disease protein. Nature Gen 1995;11:115-116.
13. Bennett, V, Stenbuck, PJ. Identification and partial purification of ankyrin, the high affinity membrane attachment site for human erythrocyte spectrin. J Biol

Chem 1979;254:2533-2541.

14. Hatzfeld, M. The armadillo family of structural proteins. Int Rev Cytol 1999;186:179-224.
15. Anastasiadis, PZ, Reynolds, AB. The p120 catenin family: complex roles in adhesion, signaling and cancer. J Cell Sci 2000;113:1319-1334.
16. Huber, AH, Nelson, WJ, Weis, WI. Three-dimensional structure of the armadillo repeat region of beta-catenin. Cell 1997;90:871-882.
17. Conti, E, Uy, M, Leighton, L, Blobel, G, Kuriyan, J. Crystallographic analysis of the recognition of a nuclear localization signal by the nuclear import factor karyopherin alpha. Cell 1998;94:193-204.
18. Riggleman, B, Wieschaus, E, Schedl, P. Molecular analysis of the armadillo locus: uniformly distributed transcripts and a protein with novel internal repeats are associated with a Drosophila segment polarity gene. Genes Dev 1998;3:96-113.
19. Kippert, F, Gerloff, DL. Highly sensitive detection of individual HEAT and ARM repeats with HHpred and COACH. PLoS ONE 2009;4:e7148.
20. Huber, AH, Weis, WI. The structure of the beta-catenin/E-cadherin complex and the molecular basis of diverse ligand recognition by beta-catenin. Cell 2001;105: 391-402.
21. Catimel, B, Teh, T, Fontes, MR, Jennings, IG, Jans, DA, Howlett, GJ, Nice, EC, Kobe, B. Biophysical characterization of interactions involving importin-alpha during nuclear import. J Biol Chem 2001;276:34189-34198.
22. Daniels, DL, Weis, WI. ICAT inhibits beta-catenin binding to Tcf/Lef-family transcription factors and the general coactivator p300 using independent structural modules. Mol Cell 2002;10: 57-584.
23. Varadamsetty, G, Tremmel, D, Hansen, S, Parmeggiani, F, Plückthun, A. Designed Armadillo repeat proteins: library generation, characterization and selection of peptide binders with high specificity. J Mol Biol 2012;424:68-87.
24. Parmeggiani, F, Pellarin, R, Larsen, AP, Varadamsetty, G, Stumpp, MT, Zerbe, O, Caflisch, A, Plückthun, A. Designed armadillo repeat proteins as general peptide-binding scaffolds: consensus design and computational optimization of the hydrophobic core. J Mol Biol 2008;376:1282-1304.
25. Alfarano, P, Varadamsetty, G, Ewald, C, Parmeggiani, F, Pellarin, R, Zerbe, O, Plückthun, A, Caflisch, A. Optimization of designed armadillo repeat proteins

by molecular dynamics simulations and NMR spectroscopy. Protein Sci 2012;21:1298-1314.

26. Nieto, JL, Rico, M, Santoro, J, Herranz, J, Bermejo, FJ. Assignment and conformation of neurotensin in aqueous solution by ^{1}H NMR. Int J Pept Protein Res 1986;28:315-323.
27. Hajduk, PJ, Meadows, RP, Fesik, SW. NMR-based screening in drug discovery. Quart Rev Biophys 1999;32:211-240.
28. Wetzel, SK, Ewald, C, Settanni, G, Jurt, S, Plückthun, A, Zerbe, O. Residue-resolved stability of full-consensus ankyrin repeat proteins probed by NMR. J Mol Biol 2010;402:241-258.
29. Cavanagh, J, Fairbrother, WJ, Palmer, AG, Rance, M, Skelton, NJ. *Protein NMR Spectroscopy: Principles and Practice* 2007(Academic Press, San Diego).
30. Sattler, M, Schleucher, J, Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog Nucl Magn Reson Spectrosc 1999;34:93-158.
31. Watson, RP, Christen, MT, Ewald, C, Bumbak, F, Reichen, C, Mihajlovic, M, Schmidt, E, Güntert, P, Caflisch, A, Plückthun, A, Zerbe, O. Spontaneous self-assembly of engineered armadillo repeat protein fragments into a folded structure. Structure 2014;22:985-995.
32. Lin, Y, Wagner, G. Efficient side-chain and backbone assignment in large proteins: application to tGCN5. J Biomol NMR 1999;15:227-239.
33. Zondlo, NJ. Aromatic-Proline Interactions: Electronically Tunable CH/π Interactions. Acc Chem Res 2012;46:1039-1049.
34. Valley, CC, Cembran, A, Perlmutter, JD, Lewis, AK, Labello, NP, Gao, J, Sachs, JN. The methionine-aromatic motif plays a unique role in stabilizing protein structure. J Biol Chem 2012;287:34979-34991.
35. Gemmecker, G. Jahnke, W, Kessler, H. Measurements of fast proton-exchange rates in isotopically labeled compounds. J Am Chem Soc 1993;115:11620-11621.
36. Reichen, C, Hansen, S, Plückthun, A. Modular peptide binding: from a comparison of natural binders to designed armadillo repeat proteins. J Struct Biol 2014;185:147-162.
37. Sambrook, J, Russell, DW. *Molecular cloning: A laboratory manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York), 2001.

38. Keller, R. The Computer Aided Resonance Assignment, 2004.
39. Jung, Y-S, Zweckstetter, M. Mars -- robust automatic backbone assignment of proteins. J Biomol NMR 2004;30:11-23.
40. Fielding, L. NMR methods for the determination of protein-ligand dissociation constants. Progr NMR Spectrosc 2007;51:219-242.
41. Christen, MT, Menon, L, Myshakina, NS, Ahn, J, Parniak, MA, Ishima, R. Structural basis of the allosteric inhibitor interaction on the HIV-1 reverse transcriptase RNase H domain. Chem Biol Drug Des 2012;80:706-716.
42. Trott, OJ., Olson, AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455-461.
43. DeLano, WL. The PyMOL Molecular Graphics Systems (DeLano Scientific, Palo Alto), 2002.
44. Schnur, J, Grieshaber, MV, Bowen, JP. Development of an internal searching algorithm for parameterization of the mm2/mm3 force-fields. J Comput Chem 1991;12:844-849.
45. Sanner, J. Python: A programming language for software integration and development. Mol Graph Model 1999;17:57-61.
46. Davis, IW, Leaver-Fay, A, Chen, VB, Block, JN, Kapral, GJ, Wang, X, Murray, LW, Arendall, WB, Snoeyink, J, Richardson, JS, Richardson, DC. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007;35:W375-W383.
47. Vriend, G. WHAT IF: a molecular modeling and drug design program. J Mol Graph 1990;8:52-6, 29.
48. Sitkoff, D, Sharp, KA, Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. J Phys Chem 1994;98, 1978-1988.
49. Berendsen, HJC, Postma, JPM, van Gunsteren, WF, DiNola, AAJRH. Molecular dynamics with coupling to an external bath. J Chem Phys 1984;81:3684-3691.
50. Bussi, G, Donadio, D, Parrinello, M. Canonical sampling through velocity rescaling. J Chem Phys 2007;126:014101.
51. Darden, T, York, D, Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. J Chem Phys 1993;98:10089-10092.
52. Best, RB, Zhu, X, Shim, J, Lopes, PE, Mittal, J, Feig, M, Mackerell, ADJ. Optimization of the additive CHARMM all-atom protein force field targeting

improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. J Chem Theory Comput 2012;8:3257-3273.

53. Jørgensen, WL, Chandrasekhar, J, Madura, JD, Impey, RW, Klein, ML. Comparison of simple potential functions for simulating liquid water. J Chem Phys 1983;79:926-935.
54. Hess, B, Bekker, H, Berendsen, HJC, Fraaije, JGEM. LINCS: A linear constraint solver for molecular simulations. J Chem Phys 1997;18:1463-1472.
55. Simons, KT, Bonneau, R, Ruczinski, I, Baker, D. Ab initio protein structure prediction of CASP III targets using ROSETTA. Proteins 1993;3:171-176.
56. Conti, E, Kuriyan, J. Crystallographic analysis of the specific yet versatile recognition of distinct nuclear localization signals by karyopherin alpha. Structure 2000;8:329-338.
57. Madhurantakam, C, Varadamsetty, G, Grütter, MG, Plückthun, A, Mittl, PR. Structure-based optimization of designed Armadillo-repeat proteins. Protein Sci 2012;21:1015-1028.

Figure Captions:

Figure 1: **Assignment strategy for fragments of increasing size.** The expansion depicts an overlay of the region of signals of Gly residues. For example, the signal of Gly254, part of the MA_{II} fragment, slightly moves towards a different position in longer fragments. In the scheme on the left the location of the peak in the smaller fragment is shown with less opacity. On the other hand, the signal of Gly86 only appears in the spectrum of the longest fragment $MR^1R^2R^3MA_{II}$.

Figure 2: Assignments of $Y_IMR^1R^2R^3MA_{II}$ in the absence (A) and presence of NT (B). Prolines and unassigned residues are colored gray; assigned Trp indole amide moieties, used to locate repeats involved in binding NT by CSP (see Figure 3B), are shown as spheres inside a red dotted cloud. **A** $Y_IMR^1R^2R^3MA_{II}$ assignment based on conventional strategy (cyan) and additional assignments achieved by fragment strategy (blue); modules are schematically labeled for reference (red). **B** $Y_IMR^1R^2R^3MA_{II}$ assignment in the presence of 2 molar equivalents of NT, using the conventional (cyan) or fragment strategy (blue).

Figure 3: CSPs induced by NT in $Y_IMR^1R^2R^3MA_{II}$ (A-C) and its truncated versions (D-F): CSP in [^{15}N,^{1}H]-HSQC spectra displaying the Gly (**A**) and indole Trp (**B**) regions of $Y_IMR^1R^2R^3MA_{II}$ in the absence (black) and presence of two molar equivalents of NT (red). Similarly, expansions of the Gly region in [^{15}N,^{1}H]-HSQC spectra of full-length $Y_IMR^1R^2R^3MA_{II}$ (**C**), $Y_IMR^1R^2R^3A_{II}$ (**D**), $Y_IMR^1R^2A_{II}$ (**E**), $Y_IMR^1A_{II}$ (**F**) are depicted with and without NT. For full spectra see Figures S4 and S5.

Figure 4: **Interaction of NT with the minimal binder $Y_IMR^1R^2A_{II}$**. **A** Backbone amide CSPs induced by addition of 2 equiv. NT are highlighted: red, resonances that could only be assigned in the presence of NT; orange, CSPs $_{HN} > \overline{\quad}_{HN} +$; yellow gradient, $_{HN} > \overline{\quad}_{HN}$; gray, Pro and unassigned residues. **B** and **C** display the two most populated binding poses obtained from MD simulations with the parallel Autodock orientation as starting point. The lowest energy ligand structure is depicted

as sticks with non-interacting residues Pro10^{NT}, Ile12^{NT} and Leu13^{NT} in gray and interacting residues in orange, blue and magenta.

Figure 5: Interaction of $Y_IMR^1R^2A_{II}$ with NT-K6C as probed by PREs. The protein backbone is in ribbon representation (gray), individually tracked atoms are shown as spheres, peptide bonds are depicted as sticks. Color intensity of the spheres is proportional to (real or simulated) PRE effect in blue-to-orange gradient. **A**, Experimental PRE effect (orange > 80% loss of signal). **B**, NTK6C-13 pose as predicted by Autodock Vina. Color intensity on the receptor correlates with distance to the nitroxide radical (orange: < 10Å ; blue: > 22Å). The docked peptide is delineated (green trace and van der Waals surface, hydrogens are hidden for clarity) with the coupled MTSL spin label (yellow trace). The nitroxide moiety (red sphere) is the source of the PRE effect as symbolized by yellow dots.

Figure 6: **Intermolecular contacts**. Time evolution of distances between Pro7^{NT}, Arg8^{NT}, Arg9^{NT}, and Tyr11^{NT} and main interacting residues of $Y_IMR^1R^2A_{II}$, when NT(7-13) is in the parallel orientation. Solid and dotted lines represent two independent runs. In (A) and (D), Pro7^{NT} - Trp77, Tyr11^{NT} – Phe126 and Tyr11^{NT}-Trp161 distances are calculated between the center of masses of the aromatic or pyrrolidine rings. In (C), Arg9^{NT} – Ser84 and Arg9^{NT} – Asn123 distances are calculated between the centers of masses of the two residues.

Figure 7: **Inter-helical angle between helix 2 of the N-cap and helix 3 of the first internal repeat**. (Left) The angle θ reflects the orientation of the N-cap relative to the first internal repeat. The black arrow indicates the position of G44 in the loop. (Middle and right) Time series and probability density, respectively, of the inter-helical angle θ for $Y_IMR^1R^2A_{II}$ (blue lines) and $Y_{III}M_3A_{II}$ (green lines).

Figure 8: **Flexibility of N-cap and first internal repeat**. (A) Profiles of RMSF of C_α atoms of selected $Y_IMR^1R^2A_{II}$ variants. (B) Profiles of RMSF of C_α atoms of $Y_IMR^1R^2A_{II}$ and $Y_IMR^1R^2A_{II}$_R37S in the presence and absence of NT. All RMSF are calculated on 2-ns segments during the second half of each simulation (i.e., between 1μs and 2μs), and then averaged. The location of helices is indicated by grey bars.

Figure 9: Temporal evolution of the distance between the S_δ atom of Met17 and the center of the phenyl ring of Phe36.

Figure 10: Workflow of the multidisciplinary approach to the elucidation of the binding mode of the NT(7-13) peptide into the designed armadillo repeat protein $Y_IMR^1R^2A_{II}$.

Figure 1

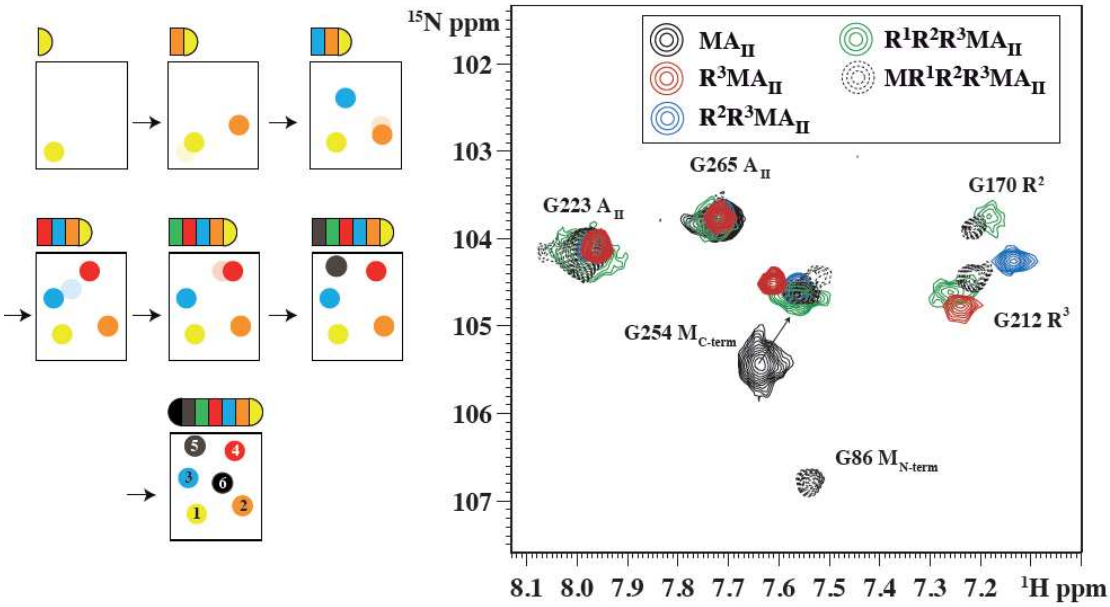


Figure 2

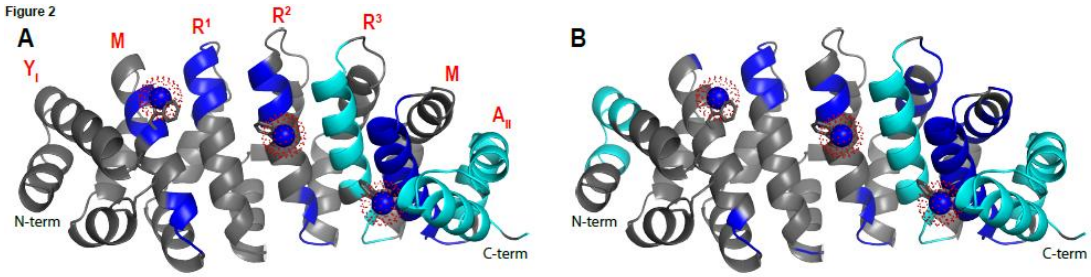


Figure 3

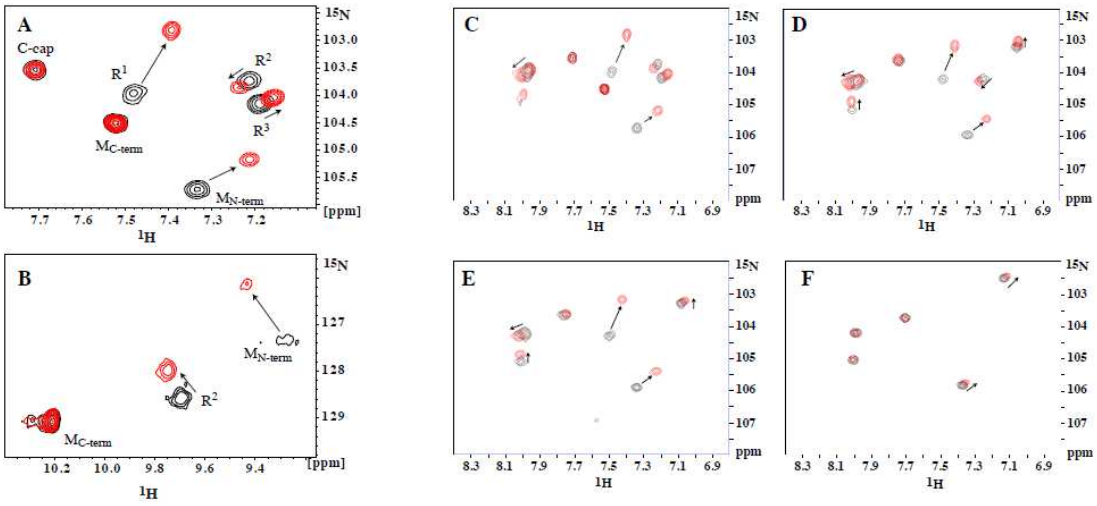


Figure 4

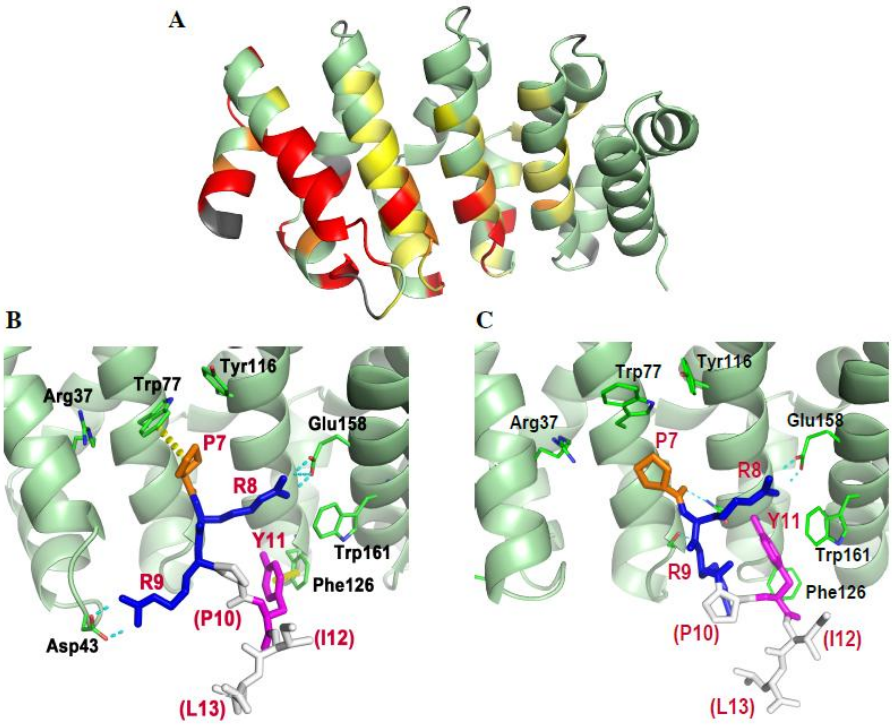


Figure 5

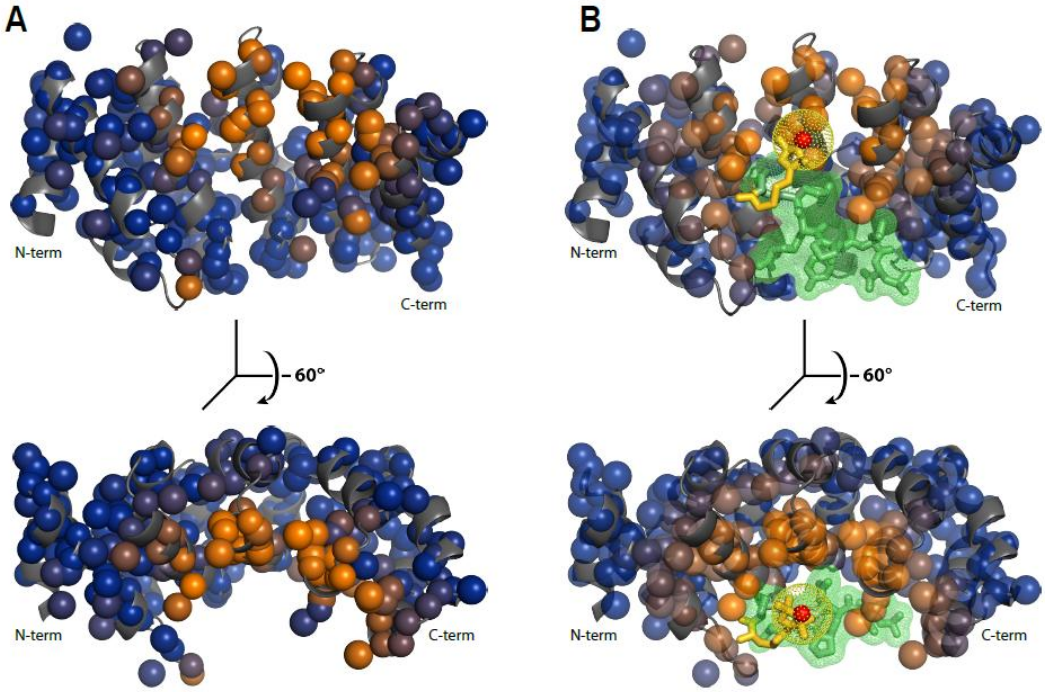


Figure 6

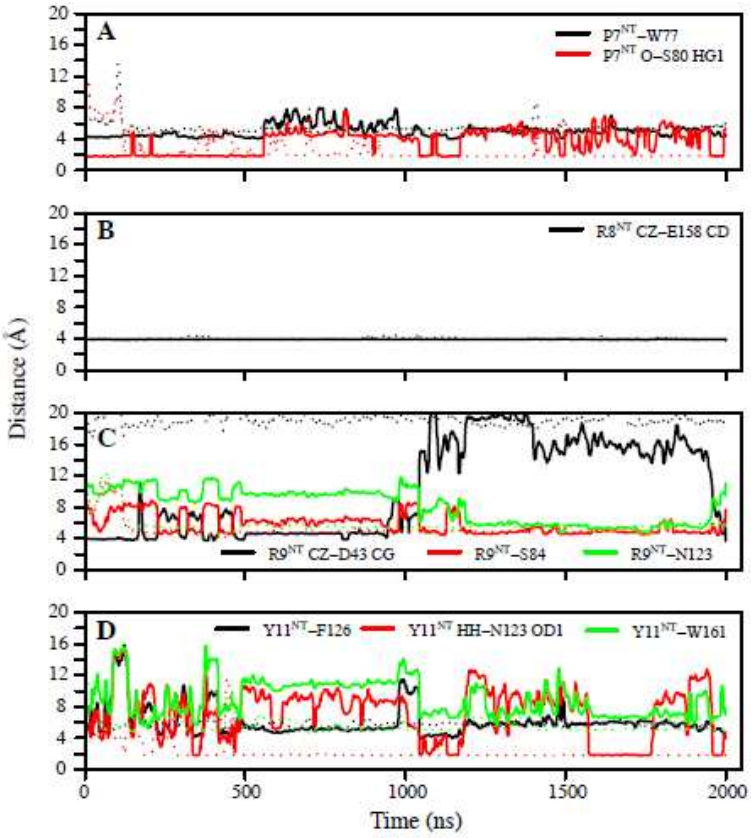


Figure 7

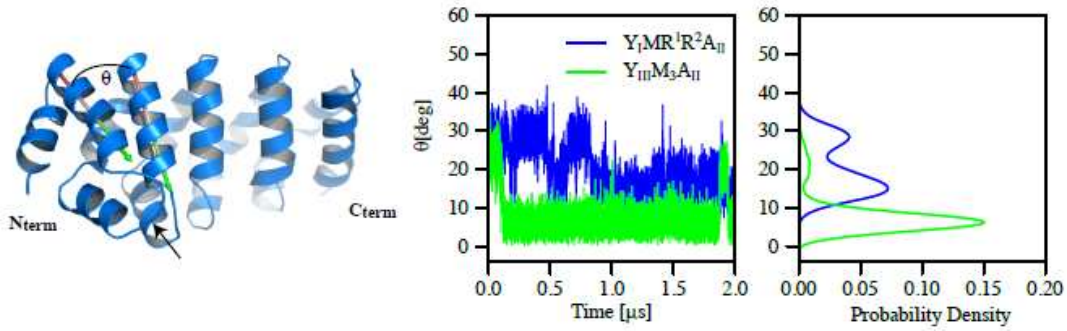


Figure 8

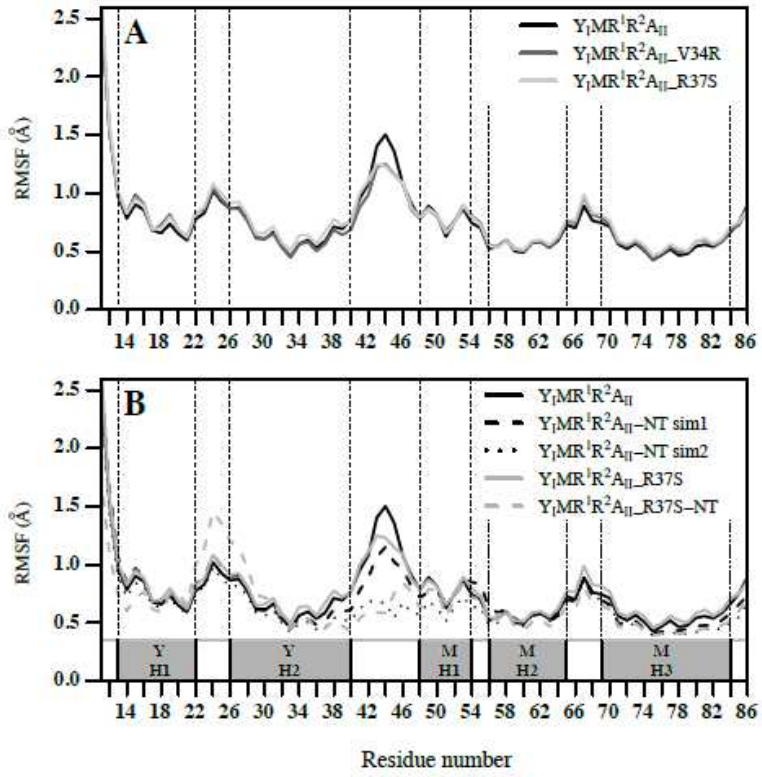


Figure 9

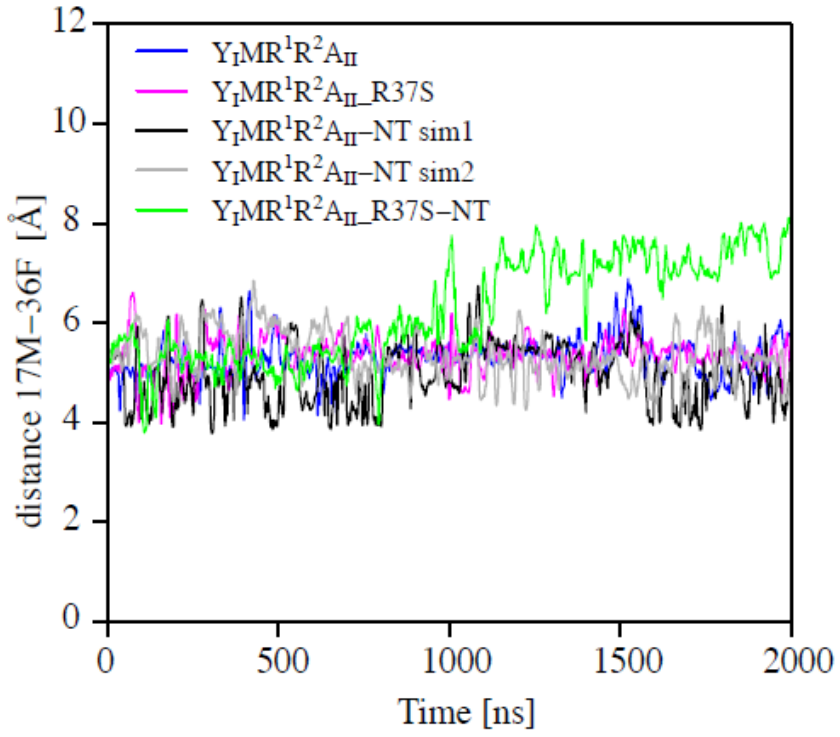


Figure 10

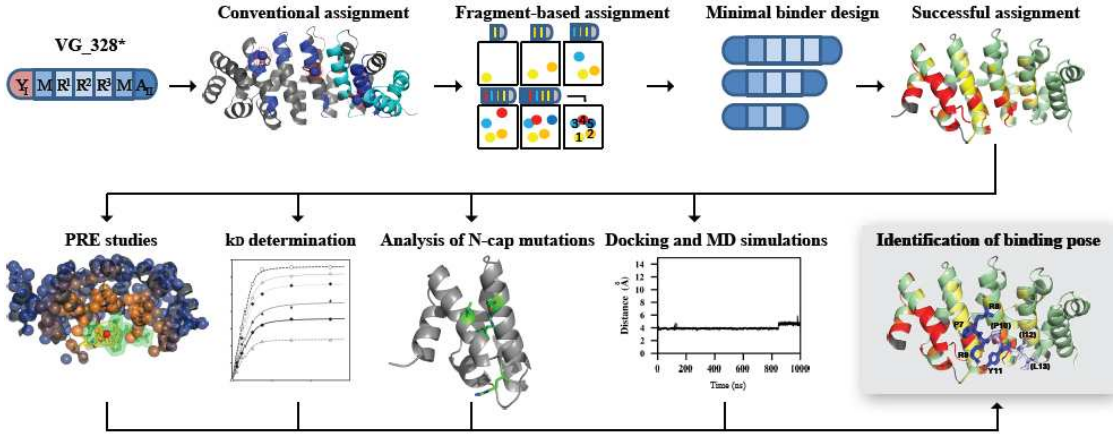
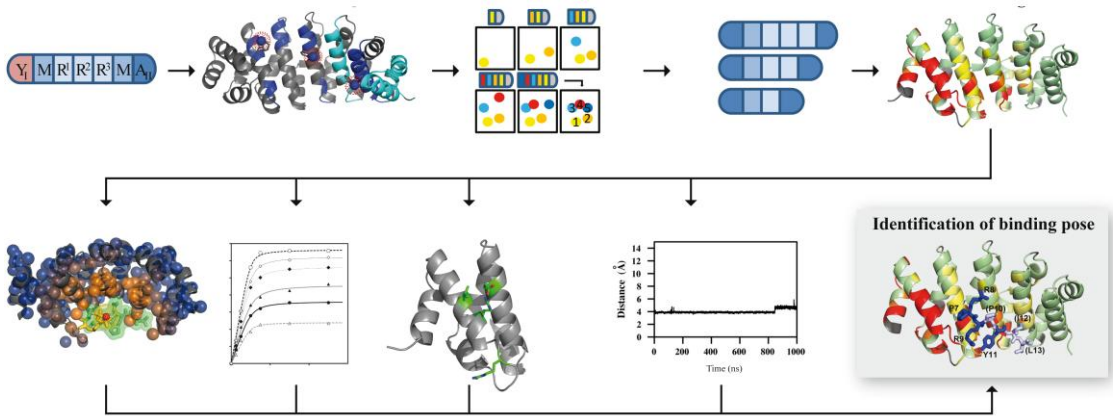


Table 1: List of dissociation constants of the complex formed between $Y_IMR^1R^2A_{II}$ mutants or VG_328-based reference proteins with NT determined by CSP titrations.

No.	Protein	Peptide	K_d [μM]	Relative K_d (relative to $Y_IMR^1R^2A_{II}$)
1	$Y_IMR^1R^2A_{II}$	NT7-13	12±5	0.7
2	$Y_IMR^1R^2A_{II}$	NT	18±7 (14 by SPR)	1.0
3	$Y_IMR^1R^2R^3MA_I$ (VG_328)	NT	20±7	1.1
4	$Y_IMR^1R^2A_{II}$_R42A	NT	25±5	1.4
5	$Y_IMR^1R^2A_{II}$_V34R	NT	27±3	1.5
6	$Y_IMR^1R^2A_{II}$_E46A	NT	27±5	1.5
7	$Y_IMR^1R^2R^3MA_{II}$	NT	27±8 (18 by SPR)	1.5
8	$Y_IMR^1R^2A_{II}$_R42Δ	NT	91±10	5.1
9	$Y_IMR^1R^2A_{II}$_V34R_R42Δ	NT	105±15	5.8
10	$Y_IMR^1R^2A_{II}$_R37S	NT	224±23	12.5
11	$Y_IMR^1R^2A_{II}$_V34R_R37S	NT	265±27	14.7
12	$Y_IMR^1R^2A_{II}$_R37S_R42Δ	NT	331±36	18.4
13	$Y_{II}MR^1R^2A_{II}$ (V34R, R37S, R42Δ)	NT	369±46	20.5



Graphical abstract

Highlights

- Determining the binding mode of weak ligands is often difficult
- A interplay of biochemical and biophysical methods enabled locating binding in repeat protein
- Preliminary chemical shift assignments indicated which repeats were involved in binding
- Subsequent elimination of repeats resulted in a smaller protein that bound the ligand equally well
- Chemical Shift Perturbations, PREs and MD data helped to locate the binding site